U.S. DEPARTMENT OF ENERGY

JOINT GENOME INSTITUTE

PROGRESS REPORT

U.S. DEPARTMENT OF ENERGY

JOINT GENOME INSTITUTE

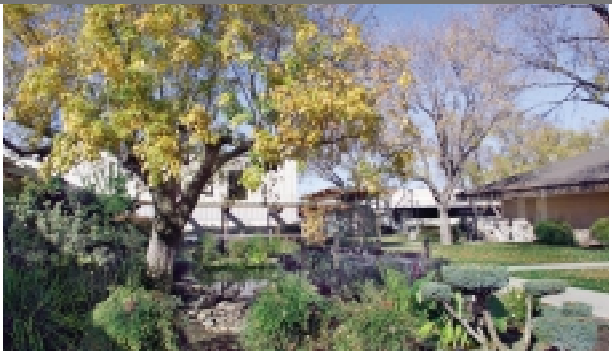
PROGRESS REPORT

JUNE 2002

". . . so we should venture on the study of every kind of animal without distaste; for each and all will reveal to us something natural and something beautiful."

Aristotle's On the Parts of Animals

DIRECTOR'S OVERVIEW



The Joint Genome Institute is transforming from a single purpose DNA sequencing facility to a genomic research institute.

The U.S. Department of Energy's Joint Genome Institute (JGI) has emerged as a leader in the international genomics effort. The detailed genetic maps and sequence information we are helping to elaborate will revolutionize medicine, agriculture, and industrial processes. Now, I am pleased to say, the JGI is transforming from a single purpose DNA sequencing facility to a genomic research institute. This Progress Report is the story of this transformation, the accomplishments of this unique institute, and our vision for JGI's future.

We have come a long way in a short time. Founded as a virtual institute on January 1, 1997, JGI presently has a 60,000 square foot Production Genomics Facility (PGF) in Walnut Creek, California, with an onsite staff of 150 and staff numbering approximately 100 at seven member and partner institutions. Original JGI members Lawrence Berkeley, Lawrence Livermore, and Los Alamos national laboratories have been joined by Oak Ridge, Pacific Northwest and Brookhaven national laboratories, and Stanford University.

In February 2001, the United States and Britain announced the accelerated completion of the working draft of the human genome. The biggest surprise coming out of this discovery was the recalculation of the number of genes from 100,000 to between 30,000 and 35,000. This observation has spurred the need for such comparative studies as the one published by JGI researchers in the July 6, 2001, issue of Science on the conserved elements of human chromosome 19 and the related region in mouse DNA. This important work confirmed evidence of genes that had been missed by other predictive methods.

JGI's achievements have not been limited to decoding the human genome. In collaboration with Baylor College of Medicine, we performed an unprecedented sequencing of an entire genome in just one day, that of the antibiotic-resistant "super bug" Enterococcus faecium, one of the leading causes of hospital-acquired infections. The information gleaned from its 2.8 million base pairs will illuminate the bacterium's vulnerabilities and allow development of more sensitive diagnostics and more effective therapies.

The glassy-winged sharp-shooter, host to Xylella fastidiosa

Building on this achievement, JGI launched Microbial Month and in October 2000 sequenced 15 different microorganisms. This came at a time when only two dozen bacterial genomes had been published. Microbial Month has yielded valuable data. Among the sequenced organisms is *Xylella fastidiosa*, a pathogen carried by an insect known as the glassy-winged sharpshooter. The pathogen infects and decimates grapevines, citrus and almond trees, and other economically important plants.

The puffer fish Fugu rubripes

Ramping off of the success of Microbial Month, JGI took on Phanerochaete chrysosporium, the white rot fungus. White rot fungi produce unique oxidative enzymes that degrade lignin, a key component of plant cell walls, as well as related compounds found in a wide variety of contaminants.

JGI's efforts also include members of the animal kingdom. Last year, we formed and led an international consortium in sequencing the puffer fish *Fugu rubripes.* The *Fugu* genome contains essentially the same genes and regulatory sequences as the human genome but packed in only 400 million bases (compared to three billion in the human genome). With far fewer noncoding sequences, the *Fugu* genome will provide a shortcut for identifying genes and important regulatory sequences in the human genome. These regulatory systems determine how or what genes are expressed—converted to structures that are actually present and operating in the cell.

The sea squirt Ciona intestinalis

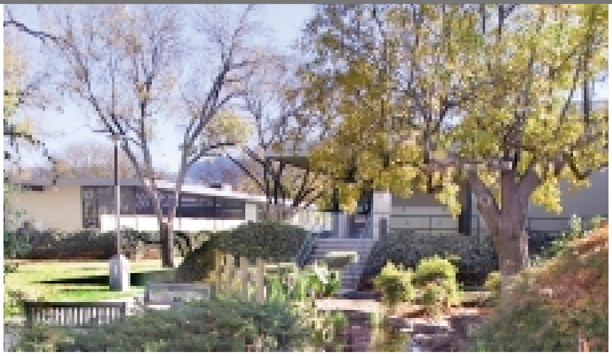
Another major sequencing project last year was the sea squirt, *Ciona intestinalis*, which provides a good system for exploring the evolution of the chordate lineage, from which all vertebrates originate. The *Ciona* genome is also the starting point of JGI's new Functional Genomics program. One of the goals of this project is to determine the regulatory networks within the simple sea squirt and find vital similarities in the human genome.

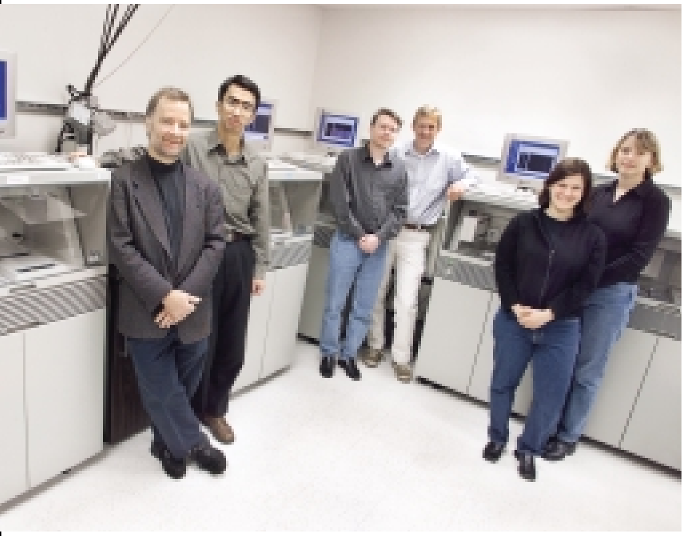
JGI continues to develop new high-throughput, genome-scale, and computational technologies. In 2001, we achieved multiple economies of scale—staff reductions, reduced costs, less space, etc.—by adapting the Rolling Circle Amplification (RCA, TempliPhi) methodology to our sequencing process. In addition, we developed a new computational algorithm called JAZZ to assemble large genome sequencing projects, such as the whole genome shotgun sequencing of the *Fugu* genome, a first in the public domain.

Our responsibility to the public is to provide free, unencumbered access to this genomic information.

The first five years of the JGI have been impressive. However, as we continue to unravel the secrets of our genes and those of other organisms, I am mindful of our role as curators of the public's trust. Our responsibility to the public is to provide free, unencumbered access to this genomic information—the critical starting material for future innovation.

JGI'S FIRST FIVE YEARS





The Joint Genome Institute was formed in 1997 as a virtual entity. The member human genome centers for the three founding national laboratory partners were managed jointly, but each laboratory had its own approaches and methods. Moreover, the initial goal for the JGI was very circumscribed—DNA sequencing of chromosomes 5, 16, and 19 of the human genome.

In 1999, the new Production Genomics Facility (PGF) in Walnut Creek, California, went on line. Although there is still a "virtual" aspect to our makeup (people are hired through the national laboratories, and sequencing and other work continue at partner and member laboratories), the PGF is the Joint Genome Institute's headquarters.

Our approaches and methods have also changed. We have transitioned from traditional sequence and finishing approaches to draft sequencing. Because our high-quality drafts contain more than 95 percent of the sequence, we can make essential data immediately available to the scientific community. A major innovation allowing this transition was the move away from the old slab gel DNA sequencer machines to the new capillary devices. JGI was the first genome center to remove all its slab gel systems and replace them with the new technology.

We are always looking for ways to scale-up our operations with new methodologies, robotics, and DNA sequencers.

The new PGF building, which had been designed to feed and run 28 of the old slab gel systems, was remodeled to accommodate 84 of the Amersham MegaBACE 1000 DNA sequencers. We have now upgraded those models to 21 of the new MegaBACE 4000 sequencers. We are always looking for ways to scale-up our operations with new methodologies, robotics, and DNA sequencers.

In pure raw base output, we produce approximately 400 bases per second, or approximately 35 million bases per day. This compares to 50 bases per second in 1997–98 and 200 bases per second as recently as 2001.

Our changes in approaches and methods and our increase in base output have allowed us to expand our horizons. From the three chromosomes of the Human Genome Project, our sequencing efforts have expanded to include organisms from all three microbial domains (bacteria, archaea, and eukarya), as well as the more familiar animal, plant, and fungal kingdoms.

THE HUMAN GENOME PROJECT

JGI has been a key player in the Human Genome Project, an international effort formally begun in October 1990 to discover all the human genes and make them accessible for further biological study. Initially believed to number around 100,000, to the surprise of all involved, the actual count will fall somewhere between 30,000 and 35,000 genes. The DOE Human Genome Program and the NIH National Human Genome Research Institute (NHGRI) together make up the U.S. Human Genome Project.

In April 2000, we announced the completion of the draft of JGI's three chromosomes (5, 16, and 19). JGI was the first of the large genome centers to make such an announcement, several months ahead of our own schedule.

The draft sequencing of the human genome has provided the 3 billion or so bases that make up the genetic information found within each of our 100 trillion cells. However, we have a poor understanding about the structure and function of the 35,000 or so genes that make up only two to three percent of our genomic sequence. In addition, we know almost nothing about the noncoding elements, which make up the rest of our genome and encode important regulatory elements such as where, when, and in what quantity proteins are made.

The three human chromosomes sequenced by the JGI contain an estimated 3,000 to 4,000 genes, including those whose defects may lead to genetically linked diseases such as certain forms of kidney disease, prostate and colorectal cancer, leukemia, hypertension, diabetes, and atherosclerosis. Knowing how these diseases are coded in our genome will allow researchers to focus on therapeutic interventions.

Ethical, Legal, and Social Issues

The Department of Energy (DOE) and the National Institutes of Health (NIH) devote three to five percent of their annual Human Genome Project budgets to the study of the project's ethical, legal, and social implications (ELSI).

People have many concerns about the project and the impact of genetic research. These concerns include:

- Predicting future illnesses before the onset of symptoms or the existence of medical therapies
- Privacy and fair use of genetic information with respect to employers, insurers, direct marketers, banks, credit raters, law enforcement agencies, and others
- Availability of genetic information in largely unprotected data banks
- Possible discriminatory misuse of genetic information

Through DOE's Human Genome Management Information System, ELSI disseminates information to the public on all aspects of genome research. For more information, visit ELSI's Web site at: www.ornl.gov/TechResources/Human_Genome/elsi/elsi.html.

Humans, mice, fish, and microbes have a lot in common. Through the process of conservation, we retain similar DNA sequences that go back millions and hundreds of millions of years to a shared ancestor.

BEYOND CHROMOSOMES 5, 16, & 19:
THE POWER OF COMPARATIVE GENOMICS

Another goal of the Human Genome Project has been to determine the complete sequence of the three billion DNA subunits of the human genome. These subunits, or nucleotides, consist of a nitrogenous base (adenine, guanine, thymine, or cytosine), a phosphate molecule, and a sugar molecule. Thousands of nucleotides are linked to form each DNA molecule.

However, it is doubtful whether the use of computational tools alone can accomplish this task. The mechanism for obtaining a full description of the human genome and understanding its structure and function is *comparative genomics*—looking at other life-forms' genetic material to interpret human gene function.

Throughout evolution, functions within cells have been defined and then not reinvented. They are instead conserved. The consequences are that genes and pathways established early in development have been maintained up to the present. Through this process of evolutionary conservation, we share a surprising number of genes with mice, fish, and even in such distant and supposedly primitive organisms as *Escherichia coli.*

Of Mice and Men

The first major example of the power of comparative genomics was published in *Science* by JGI in July 2001. This paper described the sequencing and analysis of human chromosome 19 (HSA19) and its corresponding regions in the mouse genome. It showed that sequencing of evolutionarily distinct organisms would prove to be invaluable in defining human gene structure both in the coding and noncoding regions. This noncoding area, sometimes called "junk DNA," is starting to yield valuable data.

Spanning 65–70 Mb (million bases) and containing an estimated 1,100 genes, HSA19 is one of the smallest and most gene-dense of the human chromosomes. Comparisons between HSA19 and related mouse segments identified a large number of candidate exons (the protein-coding DNA sequence of a gene) and regulatory elements associated with known genes. These analyses have also provided significant new evidence for the validity of hypothetical genes, and identified clusters of conserved sequences that offer new candidates for undiscovered HSA19 genes. Using this information, we have developed a comprehensive overview of HSA19 gene conservation and defined significant evolutionary changes that distinguish the human chromosome from related mouse DNA.

The Reader's Digest Version of the Human Genome: *Fugu rubripes*

An interesting candidate for illuminating the human genome is the Japanese puffer fish. *Fugu rubripes*, known for its poisonous enzymes, can kill an unwitting diner if the fish is prepared improperly.

"*Fugu* is just like the human genome, but at a steep discount—the Reader's Digest version," says Sydney Brenner, the two-time winner of the prestigious Lasker Award, who took up the molecular genetic study of *Fugu* a dozen years ago. Although the *Fugu* genome contains essentially the same genes and regulatory sequences as the human genome, it carries them in a package consisting of approximately 400 million bases. This is small when compared to the 3 billion bases that make up human DNA.

In November 2000 the JGI formed the Fugu Genome Consortium, which included Sydney Brenner, the Institute for Systems Biology, Myriad Genomics, Celera Genomics, and various groups in Singapore and Cambridge, UK. This has been one of the largest international consortia since the sequencing of the human genome began.

On October 26, 2001, at the 13th International Genome Sequencing and Analysis Conference in San Diego, California, the Consortium announced the completion of a draft sequence. This draft included sequencing the genome to a total of nearly sixfold coverage, and is freely available on the Web (fugu.jgi-psf.org/).

It is too early to see the full impact of the *Fugu* genome sequence, but with far less junk DNA to sort through, finding genes and controlling sequences is proving to be a much easier task. We have been able to confirm significant numbers of human genes and find many regulatory elements within the noncoding regions of the human genome.

The Searchable Squirt: Building a Platform for Discovery

Sea squirts, or *Ciona intestinalis* as they are known more formally, may look like rubbery blobs, but they are actually very advanced animals whose larvae share a similar body plan with all vertebrates, including humans. Sea squirts belong to the phylum Chordata, which includes all animals with a spinal chord and a supporting notochord (backbone). This includes everything from fish to humans.

Because *Ciona intestinalis* is perhaps the most primitive proto chordate, studying it will allow us to see how genes and gene families have evolved to higher order vertebrates. It has a very small number of cells, about 1,000 in the tadpole stage and 2,000 in the adult.

The Fugu draft represents the first vertebrate sequence assembled and made available since the Human Genome Project.

Sea squirts get their nickname from their tendency to "squirt" when they are removed from their watery home.

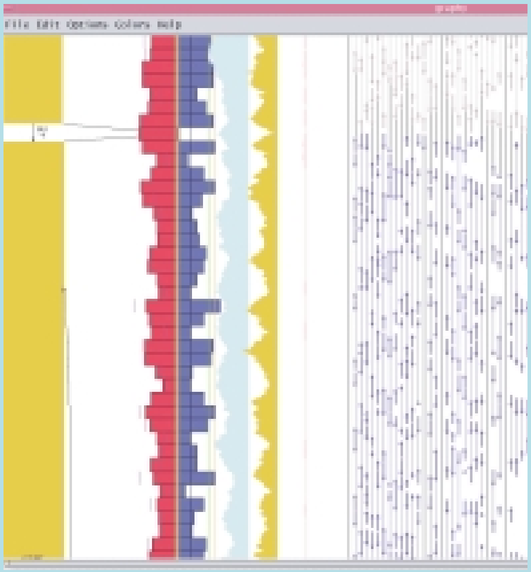
JAZZ: Piecing Together the Fragments of the Genome

One of the critical features of the Fugu project was the use of whole genome shotgun sequencing with no mapping information. To enable this approach, JGI developed the JAZZ sequence assembler program to piece the individual sequences back together. JAZZ works by assembling contiguous regions of DNA called contigs and then linking these together into regions called scaffolds.

One of the important features of JAZZ is that it allows scientists to find the "jewels" in the so-called junk of the noncoding regions. These jewels are the cis-regulatory DNA, which are part of the transcription process that determines gene expression. Unlike their protein-coding counterparts, the functions of these genes are not well understood.

JAZZ is the only sequence assembler of its type in the public domain. The development and maintenance of JAZZ is critical to the future of JGI because future projects, including the sequencing of other vertebrate and microbial genomes, will rely upon this ability.

JGI and the National Energy Research Scientific Computing Center are setting up the JAZZ software at the Center as a national resource, which will allow other centers to assemble data and make that information freely available.



On the right, each line segment represents a genomic fragment whose sequence has been determined at each end (arrows). On the left, blue rectangles represent contiguous stretches of reassembled sequence (contigs). This visualization tool allows rapid inspection of the automated assemblies produced by JAZZ. The genome shown is that of the white rot fungus Phanerochaete chrysosporium.

JGI Programs to Characterize Gene Regulatory Networks

Functional Genomics

This broad-based program will identify the noncoding regulatory networks that hide within the noncoding DNA. Using the sea squirt as a model, we will determine how networks present within the sea squirt might also be present within the human genome. Because many human diseases are caused by the misregulation of genes, we hope this work will lead to an improved awareness of the molecular mechanisms behind human health. We are also using these methods and approaches to identify regulatory networks in microbes.

Microbial Genomics

At the PGF we have the ability to sequence the genome of *E. coli* almost ten times in one day. We have used this ability to draft over 30 microbial genomes, and to date have finished five of these. These microbes have been involved in a wide range of DOE missions, including carbon sequestration, bioremediation, and human health. After the events of September 11, 2001, we have also turned our attention to the possibility of using this resource to rapidly sequence and characterize microbes that are potential biothreat agents.

Computational

This program is directed at new algorithm and tool development such as the sequence assembler JAZZ. It is also focused on analysis and annotation systems, including data dissemination and lab workflow, to enable sequence and other data samples to be tracked throughout the facility.

Dan Rokhsar, head of JGI's computational group, has described the bioinformatic tool under development as the "searchable squirt." This tool will be able to click through images of the tadpole sea squirt, identifying genes and proteins cell by cell that are part of regulatory networks, and interacting with other DNA and protein molecules.

In determining the 1,000 or so transcription factors and the roles they play in regulatory networks, we will be able to determine if these pathways and networks are present within the human genome.

The next stage will be to repeat this work with a more complex, human-like genome such as the frog. As we again identify networks in the frog, we will be able to ask if these are present in humans. In starting with the simple sea squirt and then moving on to a more complex genome, we can learn how to "read" these networks as one would read a circuit diagram. We will then have an ideal platform for discovery.

White Rot Fungus

The white rot fungus produces unique extracellular oxidative enzymes that degrade lignin, a key component of plant cell walls, as well as related compounds found in explosive contaminated materials, pesticides, and toxic wastes. To elucidate the genetic basis of this technologically important behavior, we used the JAZZ assembler to put together the 30 million base-pair genome of the white rot fungus *Phanerochaete chrysosporium,* using a whole genome shotgun method.

Phanerochaete chrysosporium has several features that might make it very useful. First, unlike some white rot fungi, when it degrades lignin it leaves the cellulose of the wood almost untouched. Second, it has a very high optimum temperature (about 40°C), which means it can grow on wood chips in compost piles, which attain a very high temperature. The exceptional oxidation potential of its enzymes has attracted considerable interest for application in bioprocesses such as organopollutant degradation and fiber bleaching. These characteristics point to some possible roles in biotechnology.

The Power of Microbes

The sequencing effort of the JGI has expanded from the Human Genome Sequencing Project to include many microbial organisms that will further DOE carbon sequestration and bioremediation missions. Three unique features allow the JGI facility to provide rapid, accurate identification of entire microbial genomes, which are then made immediately available to the scientific community:

- Exclusive Use of Paired Plasmid End Sequencing. This allows the contiguous pieces of sequence (contigs) to be ordered and oriented relative to one another.
- Sequencing Speed (One Microbe Per Day). This allows rapid targeting of genomes of interest, and the ability to attack entire related groups of microbes (e.g., the plant pathogen *Xylella*) to provide a richness of comparative information that allows genes responsible for functional differences to be identified.
- Rapid Turnaround of Assembly and Annotation. The genome of a microbe can be obtained to near completion within days and its genetic complement identified and made directly available to researchers.

The microbial sequencing effort can be directed toward microbial pathogens, including pathogens considered to be potential biothreat agents.

The white rot fungus was one of the first multicellular fungi to be sequenced.

Since its initial discovery in 1989, Vancomycin Resistant Enterococcus has become a growing concern, with the number of cases increasing twentyfold between 1989 and 1993. (The sequencing of this pathogen was a collaboration of the JGI, and George Weinstock and Barbara Murray of the University of Texas Health Science Center and Baylor College of Medicine in Houston.)

Decoding the Superbug

In May 2000 researchers at the JGI were able to complete the shotgun sequencing of "superbug" *Enterococcus faecium* in one day's worth of production time. This harmful, extremely antibiotic-resistant bacterium is one of the leading causes of hospital-acquired infections. In 1989, it was determined that strains of this bacterium have developed resistance to the "last resort" antibiotic Vancomycin.

The resistance to Vancomycin is attributed to bacterial proteins, which are coded for in the genetic makeup of *E. faecium.* Therefore, knowledge of the complete genome sequence of this organism will be crucial in developing treatments for the infections it causes. The rapid and near-universal availability of JGI's high-quality genomic library of this organism has allowed researchers worldwide to begin seeking ways to treat it. Ultimately, this rapid availability of genomic information will save lives.

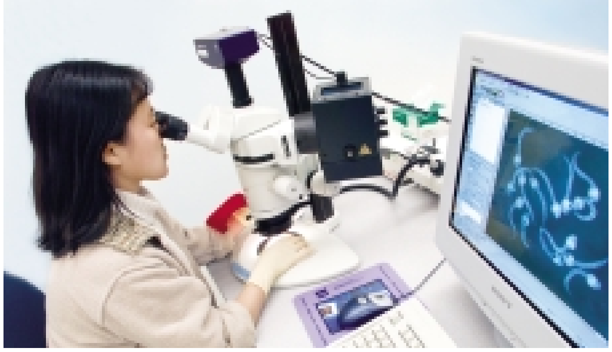
On the heels of the *Enterococcus faecium* sequencing achievement, the JGI immediately sequenced fifteen additional microbial organisms in one month, averaging one day per organism. Some of these organisms, such as *Sphingomona,* are important in bioremediation efforts. Others have a strong economic impact. For example, two strains of *Xylella fastidiosa,* which is transmitted by sharpshooters and other insects, can infect grapevines, and citrus and almond trees.

The Poplar: A Model Genome for Trees

Forest trees are the dominant life form in many ecosystems and contain more than 90% of the Earth's terrestrial biomass. Managed and unmanaged forests throughout the United States and the world provide carbon sequestration, renewable energy supplies, watershed protection, improved air quality, biodiversity, and habitat for endangered species. However, despite the importance of forest trees for natural ecosystems and the world economy, little is known about their biology when compared with the detailed information available for crop plants.

To this end, we are launching a genome sequencing initiative for *Populus.* This genus, which includes poplars, cottonwoods, and aspens, is especially well suited to serve as the model genome for trees. It has a small genome size, rapid juvenile growth, and its 30 species worldwide are easily crossed. Therefore, transgenic trees can be produced that will allow detection and characterization of gene function. *Populus* is unique among tree genera in this regard.

SCIENTISTS AND INDUSTRY COLLABORATE WITH JGI



RESEARCH COLLABORATIONS

In addition to our national laboratory partners, the Joint Genome Institute provides basic research data to the scientific community. Sequence data produced at the JGI are available free to the public on our genome portal at www.jgi.doe.gov.

JGI has also established relationships with a number of outside collaborators. As JGI transitions from being a DNA sequencing facility to a full-fledged scientific institute, we will rely more and more upon links to the scientific community to leverage expertise and knowledge. In many cases, collaborators come on sabbatical to the JGI, and we have recently established a fellowship program to allow close interaction with other scientific groups.

In early 2001, we formed the Genomic Diversity group, whose mission includes the mediation of productive collaborations between the genome science and evolutionary biology communities. This group also offers training in genome science, with an emphasis on comparative methods.

We have also developed genome jamborees to which we invite champions and experts for particular genomes that we have sequenced. These experts help us on-site with the analysis, annotation, and publication of the findings. Jamborees can last up to a week. Our first jamboree was for the white rot fungus.

Sequence data produced at JGI are available free to the public on our genome portal at www.jgi.doe.gov.

TECHNOLOGY TESTING

The JGI works with a number of companies to test and evaluate reagents and equipment at the PGF. Such arrangements are often mutually beneficial.

One technology testing success story features the TempliPhi DNA isolation methodology developed by Amersham BioSciences. This procedure uses a technique known as rolling circle amplification (RCA) to produce DNA from sub clones prior to DNA sequencing. We tested this product in its earliest forms, made suggestions, and helped define the final kit. During the testing mode, we realized that the use of TempliPhi in place of our existing methods would significantly reduce sequencing costs and personnel, and increase pass rate and read length. We began the implementation of TempliPhi in May 2001 using a prerelease version of the product, and by the time the kit was ready for full release, JGI was using it for its entire production facility. By using this product, we increased the number of bases we produce by 108%, the pass rate by 30%, and the read length by 48%.

JGI AT A GLANCE—FUNDING, STAFF, AND ORGANIZATION

The Office of Biological and Environmental Research (OBER) oversees the human genome project for the DOE and provides JGI with its DOE funding. Since January 1999, the JGI has increased its production rates more than twentyfold to approximately 35 million raw bases per day, with an essentially constant budget.

Most of JGI's staff is stationed at the Production Genomics Facility (PGF) in Walnut Creek, California. In addition, we have various groups working at the national laboratories, including a Functional Genomics group at Lawrence Livermore National Laboratory, which is responsible for the in situ hybridizations of the Ciona embryos. At Oak Ridge National Laboratory, we have staff working as part of the Computation Group. They are focused on microbial annotation and in silico analysis.

Staff are also working on finishing the drafts of human chromosomes 5, 16, and 19. At Berkeley Lab, a small group is working on mapping human chromosome 5. About 60 computational and functional genomics staff at Los Alamos National Laboratory are working on finishing human chromosome 16. About 25 full-time employees at Stanford are working on finishing human chromosomes 5 and 19.

IT GOES IN HERE AND IT COMES OUT THERE

A Sequencing Day at the Production Genomics Facility

1 Library Construction

Input: DNA from whole genomes, large clones, expressed DNA, etc.

Output: Many small clones representative of input DNA.

2 DNA Amplification and Sequencing

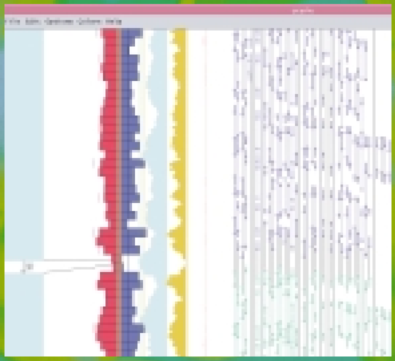
Input: Picked small clones in individual wells of a microtitre plate, grown in *E. coli* culture overnight to produce working stocks.

Output: Amplified DNA using RCA, which is then fluorescently labeled in the DNA sequencing process.

3 Electrophoresis

Input: Plates of sequenced samples in microtitre plates.

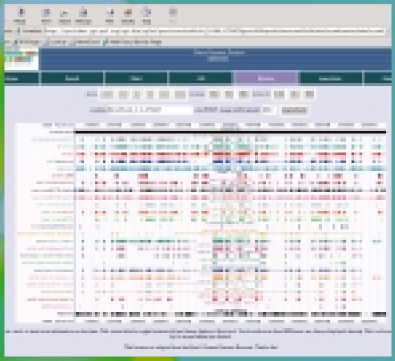
Output: Electronic file (1 read) containing chromatographic image of sequence and quality-measured base calls for every sample.



4 Sequence Assembly

Input: Thousands to millions of individual random reads for each project.

Output: Assembled genome with contiguous sequences and ordered scaffolds.

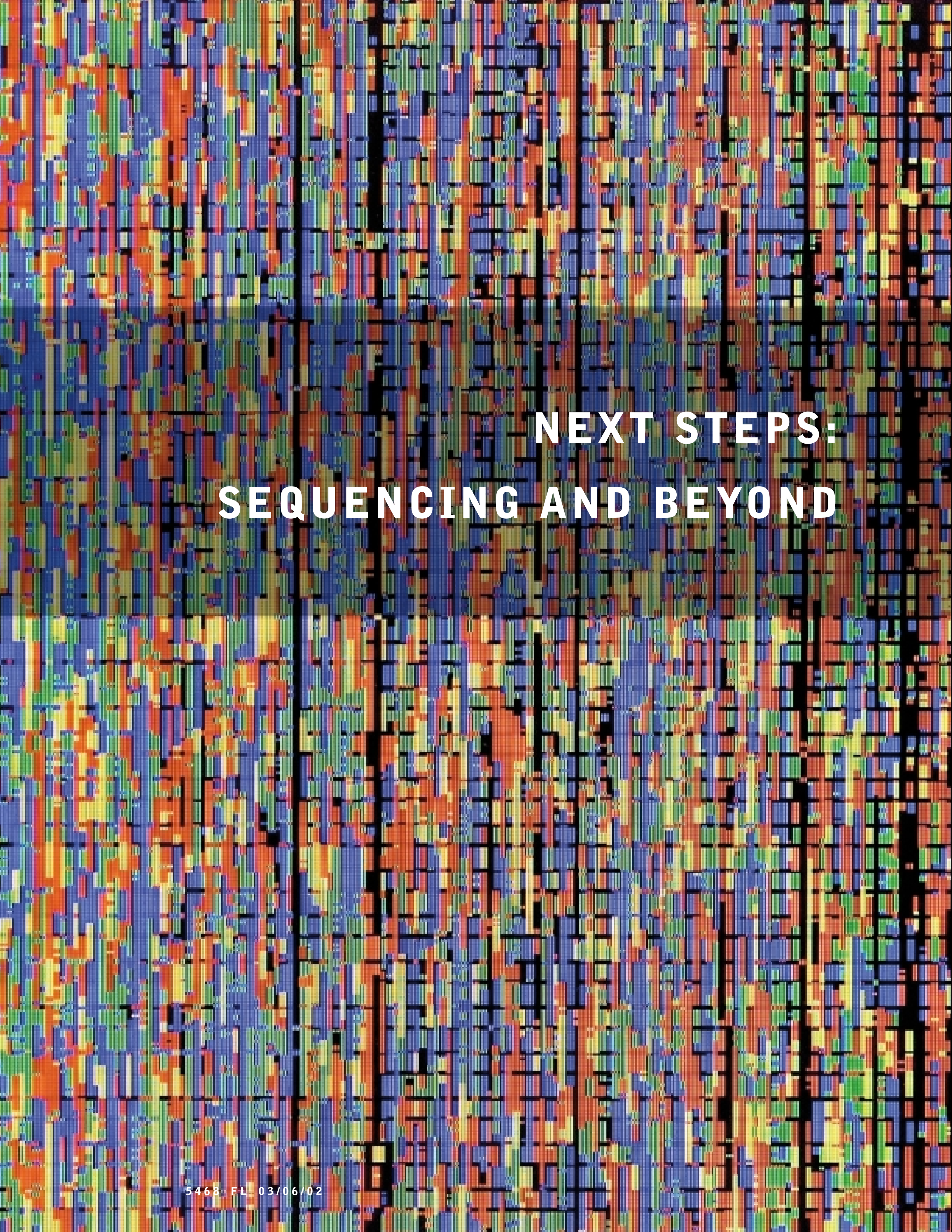


5 Data Dissemination

Input: Draft or finished genomes.

Output: Web-based tools for viewing annotated and analyzed genomic data on the JGI genome portal (at www.jgi.doe.gov).

NEXT STEPS: SEQUENCING AND BEYOND



5468_FL_03/06/02

DNA sequencing is now a commodity that we can use to answer scientific questions.

The JGI is transitioning from a DNA sequencing facility to a full-fledged genomic research institute. In that capacity, our next steps are to better understand the genes and proteins that are encoded by the human genome and to explore the vast and largely uncharted noncoding region, which makes up 97% of our DNA.

To take these next steps, we must look at other genomes, linked together by virtue of evolution, and identify common regions. These conserved elements are regions of genes and regulatory regions that are responsible for switching genes on and off. Using the human genome as a reference, we can look at other genomes, such as the puffer fish and sea squirt, to identify these conserved elements and illuminate the human genome.

However, more genomes are needed at different samplings within the genetic pool of species so that we can appreciate the structure, evolution, and function of every human gene. We can then take another significant step forward and learn how the regulation of genes is built up as a circuit and “wired” into the cells.

In preparation for this next step, we are becoming more research driven. We are not only using the sequences we produce, but are providing them—via the Web—to the public. This has led to the development of a variety of scientific programs at the JGI, including fellowship and sabbatical programs, as well as science-based collaborations of JGI and non-JGI researchers who create models from and perform analyses on the sequences we produce.

A year and a half ago, DNA sequencing was almost 100% of the JGI’s focus. Today, only 50% of our resources are dedicated to DNA sequencing, although we produce almost double the number of bases per day. DNA sequencing is now a commodity that we can use to answer scientific questions.

JGI's Microbial Genomics Program Turns Its Attention to Homeland Security

After September 11, 2001, JGI's microbial sequencing effort took on more urgency. With the ability to draft sequence a microbe in a single day, we have begun to explore the initial sequencing of a long list of well-identified biothreat agents and the rapid resequencing of pathogenic strains isolated in the field.

In close collaboration with groups at Los Alamos and Lawrence Livermore national laboratories, we are identifying ways in which this massive increase in sequence information could flow into existing programs to develop field identification and other characterization methods for pathogens. For many years it has been the role of DOE to lead the way in biothreat reduction in the U.S. JGI can bring significant and unique resources to this weighty problem.

JGI Programs Will Support DOE's Genomes to Life Initiative

The Department of Energy Office of Science's Genomes to Life (GTL) initiative has defined four specific goals aimed at gaining a fundamental understanding of the composition and function of the biochemical networks and pathways that carry out the essential processes of living organisms:

- Identify and characterize the molecular machines of life
- Characterize gene regulatory networks
- Characterize the functional repertoire of microbial communities
- Develop the computational capabilities to advance understanding of complex biological systems to predict their behavior

At the core of the GTL program is the dependence upon sequence data to drive the specific scientific goals. JGI's abilities in high-throughput/low-cost DNA sequencing will allow us to play a central role in the production of draft, finished, and annotated genome sequences for the GTL program.

June 2002

Published by the Berkeley Lab Technical and Electronic Information Department in collaboration with JGI researchers and staff for the U.S. Department of Energy.

LBNL 49907
TEID 5468